Syndrome dimensions and the Child Behavior Checklist and the Teacher Report Form: A critical empirical evaluation

Hartman, C.A.; Hox, J.; Auerbach, J.; Erol, N.; Fonseca, A.C.; Mellenbergh, G.J.; Novik, T.S.; Oosterlaan, J.; Roussos, A.C.; Shalev, R.S.; Zilber, N.; Sergeant, J.A.

published in
Journal of Child Psychology and Psychiatry
1999

DOI (link to publisher)
10.1017/S0021963099004631

document version
Publisher's PDF, also known as Version of record

J. Child Psychol. Psychiat. Vol. 40, No. 7, pp. 1095–1116, 1999
Cambridge University Press

0021–9630/99 $15.00+0.00

Syndrome Dimensions of the Child Behavior Checklist and the Teacher Report Form: A Critical Empirical Evaluation

Catharina A. Hartman
University of Amsterdam, The Netherlands

Joop Hox
University of Amsterdam, The Netherlands

Judith Auerbach
Ben Gurion University, Beer Sheva, Israel

Nese Erol
University of Ankara, Turkey

Antonio C. Fonseca
University of Coimbra, Portugal

Gideon J. Mellenbergh
University of Amsterdam, The Netherlands

Torunn S. Nøvik
National Centre for Child and Adolescent Psychiatry, Oslo, Norway

Jaap Oosterlaan
University of Amsterdam, The Netherlands

Alexandra C. Roussos
Attiki Child Psychiatry Hospital, Greece

Ruth S. Shalev
Shaare Zedek Medical Centre, Jerusalem, Israel

Nelly Zilber
JDC Falk Institute for Mental Health and Behavioural Studies, Jerusalem, Israel

Joseph A. Sergeant
University of Amsterdam, The Netherlands

The construct representation of the cross-informant model of the Child Behavior Checklist (CBCL) and the Teacher Report Form (TRF) was evaluated using confirmatory factor analysis. Samples were collected in seven different countries. The results are based on 13,226 parent ratings and 8893 teacher ratings. The adequacy of fit for the cross-informant model was established on the basis of three approaches: conventional rules of fit, simulation, and comparison with other models. The results indicated that the cross-informant model fits these data poorly. These results were consistent across countries, informants, and both clinical and population samples. Since inadequate empirical support for the cross-informant syndromes and their differentiation was found, the construct validity of these syndrome dimensions is questioned.

Keywords: Child behaviour, classification, concept of development, psychometrics, symptomatology, confirmatory factor analysis.

Abbreviations: ADF: Asymptotic Distribution Free; ADHD: Attention Deficit/Hyperactivity Disorder; CBCL: Child Behavior Checklist; CFA: confirmatory factor analysis; CFI: Comparative Fit Index; CTRS: Conners Teacher Rating Scale; EFA: exploratory factor analysis; GFI: Goodness of Fit Index; HKD: Hyperkinetic Disorder; ML: Maximum Likelihood; MTMM: Multitrait-Multimethod matrix; PCA: Principal Component Analysis; PMCC: product moment correlation coefficient; RMR: Root Mean Square Residual; RMSEA: Root Mean Square Error of Approximation; TRF: Teacher Report Form; ULS: Unweighted Least Squares; YSR: Youth Self Report.

Requests for reprints to: Catharina A. Hartman, University of Groningen, Faculty of Medical Sciences, Department of Psychiatry, Hanzeplein 1, Entrance 32, PO Box 30001, 9700 RB Groningen, The Netherlands (E-mail: c.a.hartman@med.rug.nl).
Catharina A. Hartman is currently working at the Faculty of Medical Sciences, University of Groningen, The Netherlands.

In both the clinical-diagnostic tradition and the empirical-quantitative tradition, taxonomies of childhood psychopathology have developed in recent years from relatively undifferentiated to specific concepts (Achenbach, 1995; Cantwell, 1996; Volkmar & Schwab-Stone, 1996). However, in the clinical-diagnostic tradition, both the diagnostic labels and the criteria used for the clinical assessment of childhood psychiatric syndromes have been the subject of debate (see DSM; American Psychiatric Association, 1980, 1987, 1994; and ICD; World Health Organisation, 1967, 1978, 1992). Clinically derived taxonomies have been criticised further for their lack of empirical support (Achenbach, 1995; Quay, 1986a, b). In contrast, the quantitative-empirical approach to conceptualising childhood psychiatric syndromes has provided heterogeneous findings with regards to which symptoms measure which problem dimensions. That is, despite some consistency of global clusters across empirical studies (Quay, 1986b), delineation of these childhood syndrome dimensions is still imprecise. Critics of the quantitative-empirical approach suggest that there is little congruity with regards to both the number and the nature of problem dimensions that are mutually necessary and sufficient to represent various domains of psychopathology (Millon, 1991). In short, to date there is neither agreement nor empirical evidence regarding exact operationalisation of childhood psychiatric syndromes.

Consequently, instruments with apparently comparable coverage differ with regard to which syndrome dimensions are indexed by which symptoms. For example, the modified Conners Teacher Rating Scale (CTRS-28; Goyette, Conners, & Ulrich, 1978) and the Teacher Report Form (TRF; Achenbach, 1991b) contain respectively the dimensions of inattentive-passive and attention problems. Both of these empirically derived instruments address the construct "inattention". They differ, however, in that the TRF "attention problems" scale contains items such as "can't sit still", "impulsive", and "fidgets", which are elements of the hyperactive/impulsive domain of the clinical diagnosis Attention-Deficit/Hyperactivity Disorder (ADHD in DSM-IV; American Psychiatric Association, 1994) or Hyperkinetic Disorder (HKD in ICD-10; World Health Organisation, 1992) (for a discussion of ADHD and HKD, see Swanson et al., 1998). In contrast, the CTRS distinguishes between the dimensions "inattention" and "hyperactivity". Furthermore, the TRF contains in the "aggression" scale items that are part of the CTRS's "hyperactivity" scale: "disturbs other children", "demands must be met immediately", and "demands a lot of attention". Thus, the constructs comprised in the TRF and the CTRS, while containing apparently similar problem dimensions, differ in content and domain, and differ as to how they are related to the clinical diagnosis of ADHD/HKD. In short, a typology of childhood psychiatric syndromes, whether originating from the clinical or empirical tradition, is in a state of "work in progress".

Part of this process is the sharpening of definitions and criteria. As Quay (1986a, p. 2) put it: "We can never arrive at a scientific understanding of any specific disorder until we can describe it accurately and determine how it is different from other disorders". Research effort has been concerned primarily with criterion-related validation through aetiological, prognostic, or treatment outcome studies (Frick et al., 1994; Lahey, Applegate, Barkley et al., 1994; Lahey, Applegate, McBurnett et al., 1994). However, the power of these external construct validation studies depends upon the adequacy with which the diagnostic groups are defined and selected. This in turn depends upon the conceptual coherence of the symptoms in syndromes and the precision with which these can be differentiated from one another. Ideally, clarification of the internal construct validity of diagnostic syndromes and their defining criteria should occur prior to validation with respect to external criteria (Waldman, Lilienfeld, & Lahey, 1995).

The present paper is concerned with the construct representation of the empirically derived syndrome dimensions used in the CBCL (Achenbach, 1991a) and the TRF (Achenbach, 1991b). The work of Achenbach and associates is one of the major efforts towards a quantitative empirically defined taxonomy of childhood psychopathology. Furthermore, this research is an excellent example of how the issuwe of combining information from multiple informants may be addressed. In addition, this research programme included attempts to ascertain the cultural (in)dependence of the empirical taxonomy (Berg, Fombonne, McGuire, & Verhulst, 1997; De Groot, Koot, & Verhulst, 1994, 1996). These instruments have been translated into 55 languages. Over 2300 publications report both practical and research applications of these instruments. Given their widespread use, a thorough investigation of the construct representation of the cross-informant syndromes seems warranted.

In the present paper, the eight cross-informant syndromes of the CBCL and the TRF are studied in samples collected in seven different countries, separately for parents and teachers. The objective is to determine the internal construct validity of these syndrome dimensions. A strong test of the construct validity of the syndrome representation of an instrument is the replication of its factor structure in different cultures (Bird, 1996; Verhulst, 1995). Furthermore, empirical support for the validity of similar syndrome dimensions across informants is a prerequisite for uniformity of measurement instruments. A factor analytic approach was used in the present study to determine internal construct validity, as described below.

The Factor Analytic Approach

In this study, construct validity is investigated within a factor analytic framework. A factor is a latent variable on which individuals vary. In factor analysis the underlying constructs are assumed to be continuous. The factor analytic model provides a dimensional view of childhood syndromes. In the dimensional tradition, differences between children's scores on a particular syndrome dimension are viewed as quantitative, varying in intensity rather than in quality. "Normal" children will have certain scores on the dimension and children who have certain problems will have other scores on the dimension. Thus, even when the child is typically *not* an aggressive person, the construct "aggression" is still relevant to him/her (see Jackson, 1973). It is further assumed that there is no discontinuity between those children who have the syndrome and those who do not have the syndrome. The eight syndromes of the CBCL and the TRF were derived in this dimensional tradition (Achenbach, 1991a, b).

In the factor analytic framework the latent variables are viewed as theoretical abstractions that cannot be observed directly. They can, however, be assessed by considering the degree to which the associated observable variables are present. This distinction between the unobservable and the observable is also fundamental to developmental psychopathology (Rutter & Pickles, 1990). Different types of psychopathology are regarded as referring to different constellations of symptoms. The symptoms are regarded as representative but imperfect indices of the syndrome. They are the basis from which the presence of underlying unobservable syndrome of the child is deduced. Thus, both factor analysis and child psychopathology assume latent underlying constructs which have measurable attributes.

Factor analysis assesses the construct validity of syndrome dimensions by examining the internal structure of the instrument through modelling the patterns of covariation among the measurable attributes. The notion that the presence of certain symptoms is a manifestation of a particular underlying syndrome implies that these symptoms occur together to some extent in children with that syndrome. Symptoms that are not features of the syndrome are less likely to be present in these children. Children with another syndrome are, in turn, more likely to exhibit the symptoms that are regarded as diagnostic indicators for this second syndrome. Children with no syndromes are likely to have low scores on the symptoms throughout. Syndrome dimensions are thus implied by the patterns of covariance among the symptoms. Not all the symptoms need to be present or absent to the same degree, but the more the hypothesised pattern is present, the more coherent and differentiated will be the underlying hypothesised problem dimensions (Muthén, Hasin, & Wisnicki, 1993). A good match between the pattern of covariation predicted by the factor model and that observed in the data suggests that there is empirical support for the hypothesised model representing the syndrome dimensions, the items by which these syndrome dimensions are indexed, and their differentiation.

Two approaches can be distinguished in factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). These approaches differ in the degree of a priori hypothesised explicitness of the patterns of clustering of the symptoms. EFA emphasises the exploration and identification of the latent variables and their indicators. The syndrome dimensions of the CBCL and the TRF were developed in this inductive manner by means of Principal Component Analysis (PCA). This method is similar to EFA in that both techniques seek to identify underlying dimensions of observed variables. Items were chosen such that they formed a representative and wide range of child psychiatric problem behaviours from which the syndrome dimensions were empirically derived (Achenbach, 1991a, b). CFA, on the other hand, aims at the confirmation of a hypothesised factor structure rather than exploration. The explicit factor structure derived in an exploratory manner for the CBCL and the TRF (Achenbach 1991a, b), may be followed up with more formal hypothesis testing with CFA in subsequent samples.

A schematic representation of the cross-informant factor model, as well as a list of the relevant problem items for both type of informants, is presented in Fig. 1. Figure 1 is a representation of the general factor analytic equation:

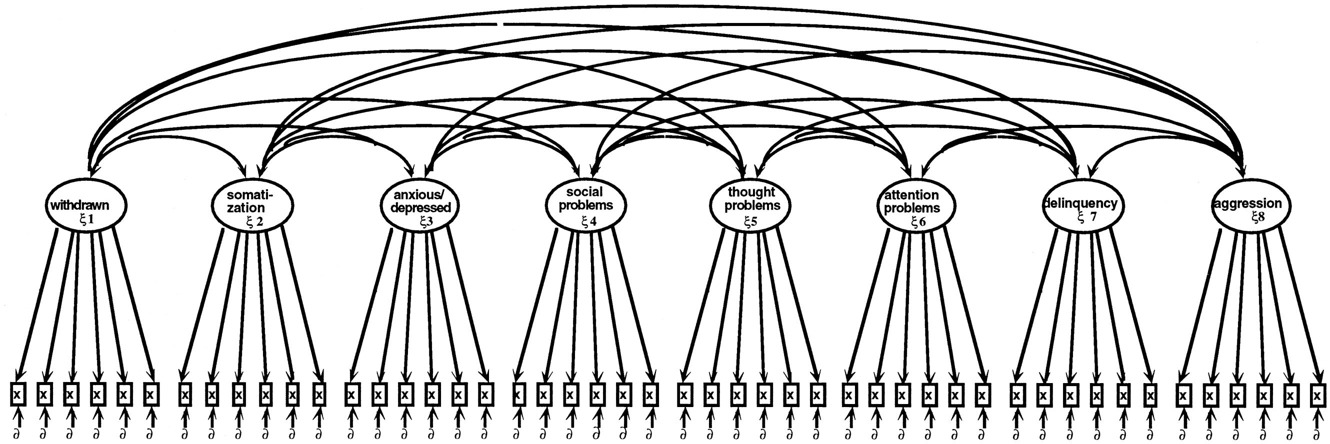
$$\mathbf{x} = \Lambda_{\mathbf{x}} * \xi + \delta,$$

where $\mathbf{x}$ is the vector of observed variables, i.e. the problem items; $\Lambda_{\mathbf{x}}$ is a matrix of factor loadings; ξ is a vector of common factors; and δ is the vector of specific factors. Thus, there are two categories of latent variables: common factors and specific factors. The common factors represent the underlying syndrome dimensions, which give rise to the covariation between the problem items. The specific factors are responsible for variation unique to each problem item. In short, a child's score on a problem item is determined partly by the child's score on the syndrome dimension specified by the model, and partly by unique variance.

The model shown in Fig. 1 is a confirmatory factor model. Instead of all problem items loading on all underlying constructs, as in an exploratory factor model, the measurement structure is defined by a specific prespecified pattern of items loading on specific constructs. In Fig. 1 it is indicated that the items in the model follow a simple structure, i.e. a child's score on a particular problem item is dependent on only one underlying syndrome dimension in themodel (see Jöreskog, 1979a). Figure 1 is simplified in two respects: first, the number of items is not constant but varies for each of the eight cross-informant syndromes. Second, a small number of items, as specified in the cross-informant model, load on more than one syndrome dimension and are thus of complex structure (see legend in Fig. 1).

Additional model specifications may be derived from Fig. 1. First, the covariations among the specific factors are required to be zero in order not to introduce additional symptom covariation over and above the eight hypothesised cross-informant syndromes. Second, the double-headed arrows connecting the common factors indicate that the common latent constructs covary. Although originally the syndrome dimensions were derived using an orthogonal rotation procedure (Achenbach, 1991a, b), the requirement of uncorrelated problem dimensions seems to be too stringent, given the syndrome overlap in child psychopathology (Angold, Costello, & Erkanli, 1999; Caron & Rutter, 1991; Sonuga-Barke, 1998). A correlated factor model (or an oblique rotation) may thus be a more realistic choice, resulting in factors that potentially have a better chance to be integrated in existing theory. The use of a correlated factor model is consistent with previous CFA studies on the measurement structure of the CBCL (Berg et al., 1997; Dedrick, Greenbaum, Friedman, Wetherington, & Knoff, 1997; De Groot et al., 1994; Van den Oord, 1993) and the TRF (De Groot et al., 1996).

The goodness of fit of a factor model is indicated by the degree to which the theoretical covariance (or correlation) structure implied by the hypothesised cross-informant



WITHDRAWN

common
42 Rather BeAlone
65 Refuses Talk
69 Secretive
75 Shy
80 Stares*
88 Sulks
102 Underactive
103 Sad*
111 Withdrawn

SOMATIC COMPLAINTS

common
51 Dizzy
54 Overtired
56a Aches
56b Headaches
56c Nausea
56d Eye
56e Skin
56f Stomach
56g Vomit

ANXIOUS/DEPRESSED

common
12 Lonely
14 Cries
31 FearDo Bad
32 BePerfect
33 Unloved
34 Persecuted
35 Worthless
45 Nervous*
50 Fearful
52 Guilty
71 SelfConscious
89 Suspicious
103 Sad*
112 Worries

only TRF
47 Overconforms
81 HurtCriticized
106 Anxious ToPlease
108 Afraid Mistakes

SOCIAL PROBLEMS

common
1 ActsToo Young*
11 Clings
25 NotGet Along
38 Teased
48 NotLiked
62 Clumsy*
64 Prefers Younger

only CBCL
55 Overweight

only TRF
12 Lonely#
14 Cries#
33 Unloved#
34 Persecuted#
35 Worthless#
36 Accident Prone

THOUGHT PROBLEMS

common
9 Can'tGet MindOff Thoughts
40 Hears Things
66 Repeats Acts
70 Sees Things
84 Strange Behavior
85 Strange Ideas

only CBCL
80 Stares#

only TRF
18 HarmsSelf
29 Fears

ATTENTION PROBLEMS

common
1 ActsToo Young*
8 Can'tConcentrate
10 Can'tSit Still
13 Confused
17 DayDreams
41 Impulsive
45 Nervous*
61 Poor School
62 Clumsy*
80 Stares*

only CBCL
46 Twitches

only TRF
2 Hums
4 FailsTo Finish
15 Fidgets
22 Directions
49 Learning
60 Apathetic
72 Messy
78 Inattentive
92 Underachieving
100 CarryOut Tasks

DELINQUENT BEHAVIOR

common
26 NoGuilt
39 BadCompanions
43 LieCheat
63 Prefers Older
82 Steals Outside
90 Swears
101 Truancy
105 Alcohol Drugs

only CBCL
67 Runs Away
72 SetsFires
81 StealsAt Home
96 Thinks Sex
106 Vandalism

only TRF
98 Tardy

AGGRESSIVE BEHAVIOR

common
3 Argues
7 Brags
16 CruelMean
19 Demands Attention
20 Destroys Own
21 Destroys Other
23 Disobed. School
27 Jealous
37 Fights
57 Attacks
68 Screams
74 ShowOff
86 Stubborn
87 Mood Changes
93 TalksToo Much
94 Teases
95 Temper
97 Threaten
104 Loud

only CBCL
22 Disobed. Home

only TRF
6 Defiant
24 Disturbs
53 TalksOut OfTurn
67 Disrupts class
76 Explosive
77 Frustrated

Figure 1. Schematic representation of the CBCL and TRF cross-informant measurement structure and list of pertaining problem items (Model 6). The problem item numbers correspond with the numbers in the CBCL and the TRF; symptom content is summarised; the common item model is based on the 77 items common to the CBCL and the TRF cross-informant model; the full cross-informant model is based on 85 symptoms for the CBCL, and 101 for TRF; * indicates that the symptom is of complex structure in the common symptom cross-informant model; # indicates that symptom is of complex structure in the full cross-informant model.

model and the observed sample covariance (or correlation) matrix agree with one another. When this fit is found to be acceptable for the parent data, as well as the teacher data, the eight syndrome dimensions are judged to provide an adequate summary of the covariation patterns among the problem items. This would provide support for the internal construct validity of the cross-informant model.

Having established that the fit is acceptable, the correlations among the factors and the factor loadings may be interpreted. This provides information on the relative independence of each of the syndrome dimension, as well as the degree to which each of the problem items is a central feature of the syndrome dimension.

Confirmatory factor analysis has been applied in a number of studies using the CBCL (Berg et al., 1997; Dedrick et al., 1997; De Groot et al., 1994; Van den Oord, 1993), and in one study using the TRF (De Groot et al., 1996). Previous studies restricted their analysis to only one method: Unweighted Least Squares (ULS) applied to polychoric correlations (described below), using conventional rules to assess goodness of fit. In the present study two methods are used: Maximum Likelihood (ML) applied to product moment correlation coefficients and ULS applied to polychoric correlations. These methods are complementary: ML is the most commonly method used in CFA (see, e.g. Marsh, Hau, Balla, & Grayson, 1998), and product moment correlation coefficients (PMCCs) are relatively stable. In contrast, ULS can be applied to matrices that are deficient in rank (Wothke, 1993), which is the case for the polychoric correlation matrices in our samples. Polychoric correlations, although more unstable, may provide more accurate estimates of the underlying associations between the symptoms. These methods are described in more detail in the Method section of this paper. The adequacy of fit for the cross-informant model is established here on the basis of three approaches: conventional rules of fit, simulation, and comparison with other models (also described below). In addition, the diversity and volume of the samples reported here are unequalled. The central question is: Is there sufficient evidence for the factorial validity of the empirically defined taxonomy of the CBCL and TRF to justify its use and interpretation?

Method

Subjects

Data were collected from the following seven countries: Greece, Israel, Norway, Portugal, the Netherlands, Turkey, and the United States of America. Table 1 lists the age and gender distributions of each of the samples. These samples have been described in detail elsewhere. The Norwegian (Nøvik & Zeiner, 1995), Turkish (Erol, Arslan, & Akçakin, 1995), Portuguese (Fonseca et al., 1995), Israeli (Zilber, Auerbach, & Lerner, 1994), Greek (Hartman et al., 1995), and United States (Loeber, Farrington, Stouthamer-Loeber, & Van Kammen, 1998) data were sampled from the general population. Two samples consisted of clinically referred children, a Dutch and an Israeli sample (Zilber et al., 1994). The Israeli teacher sample was a mixed sample (Auerbach, Goldstein, & Elbadour, 1998). About half of this sample was rated by the teachers as having problems to the extent that clinical evaluation was warranted. For the Norwegian subjects only CBCL ratings were available. Both CBCLs and TRFs were available for the Dutch, Greek, Israeli, Portuguese, Turkish, and US samples. Each sample was analysed separately, since we did not want to assume a priori that the samples are homogeneous, i.e. they can be described by the same model (see, e.g., Muthén, 1989). Samples that are different with respect to some known external criterion (e.g. country) may have a different factor structure. Likewise, we did not model parent and teacher data simultaneously in a Multitrait-Multimethod matrix (MTMM) (Campbell & Fiske, 1959; see, for an example of CFA applied to MTMM data, Kenny & Kashi, 1992). This procedure averages out rather than illuminates potential differences between parent and teacher populations (see Wothke, 1996, for more complex statistical models than CFA which allow for *interactions* between traits and methods).

Cross-informant Model

The CBCL and the TRF are questionnaires for assessing problem behaviours and competencies of children as reported by their parents and teachers, respectively (Achenbach 1991a, b). The part of these instruments relating to problem behaviour consists of 120 problem items. These items are rated using a 3-point Likert scale, where 0 indicates responses of "not true", 1 "somewhat or sometimes true", and 2 "very true or often true". Achenbach developed a cross-informant model that is similar for both sexes, a large age range (4 to 18 years), and for three informants (parent, teacher, child). This model consists of 8 syndromes, which are measured by 85 of the 120 items for the CBCL and 101 of the 120 items for the TRF (see Achenbach, 1991a, p. 45; 1991b, p. 36, for an exact representation of the cross-informant model), which is given in Fig. 1 here. Consequently, the eight syndromes are partly indexed by different items for parents and teachers. Seventy-seven problem items of the cross-informant syndromes are common to both instruments.

Both the full cross-informant model, based on 85 and 101 items, respectively, and the restricted cross-informant model, based on the 77 common items, were fitted to the data. These models are referred to as the full cross-informant model and the common item model, respectively. The separation of these two models facilitates comparisons of model fit of parents and teachers, since for the common item model only the informants reporting the problems of the child differ, but the problem items themselves do not.

Data Analysis

Child psychiatric symptoms do not fulfil the factor analytic requirements of normally distributed variables. They are known

Table 1
Sample Characteristics

	Greece		Portugal		Turkey		Norway	Netherlands		Israel			United States	
Sample	POP		POP		POP		POP	CLR		CLR	POP	MIX	POP	
Age range	6–12		6–16		6–18		4–17	4–18		4–17	4–17	6–11	5–16	
Source	CBCL	TRF	CBCL	TRF	CBCL	TRF	CBCL	CBCL	TRF	CBCL	CBCL	TRF	CBCL	TRF
Total	1213	1179	1375	1377	1564	1608	1162	1753	1418	2246	1340	954	2573	2357
Boys	602	581	700	719	752	792	570	1174	955	1384	672	539	2573	2357
Girls	611	598	675	658	812	816	592	579	463	862	668	415	–	–

POP: population sample; CLR: clinically referred sample; MIX: mixed sample.

to be skewed (see, for example, Farrington & Loeber, in press). Furthermore, as with most questionnaires in child psychiatry, the CBCL and the TRF are scored on an ordinal (3-point) rather than a continuous scale. There is no agreed best method for factor analysing a large number of highly skewed, ordinally scored items with restricted sample sizes. Due to space limitations a brief account of the relevant problems and the choices made with regard to the data analysis is provided here.

Measure of association. The first choice to be made for the data here concerns the type of measure of association to be used. Given that the data are categorically measured, a polychoric correlation (Jöreskog, 1994; Olsson, 1979), rather than a covariance or PMCC would seem to be the best choice. The reason for this is that the maximum value of the covariance (or correlation) between two categorically scored items is often downwardly biased (Farrington & Loeber, in press; Muthén, 1989). This bias increases when the number of response categories is small, and as the item responses depart from equal representation in the response categories. These attenuated correlations result in downwardly biased factor loadings. This would incorrectly indicate poor reliability and validity of the items. Simulation studies have generally shown that polychoric correlations do not suffer from this problem and that they provide accurate estimates of pairwise correlations (Babakus, Ferguson, & Jöreskog, 1987).

However, some properties of the data in the present study argue against the use of polychoric correlations. First, the assumption of underlying bivariate normality of the variables required for polychoric correlations may be unrealistic. It is improbable that the skewness of the item scores can be attributed solely to crude measurement. Even use of a continuous scale would reveal that the majority of children cluster in the "no problem" range. Second, the considerable skew creates a paucity of observations in the 1 and 2 response categories and the bivariate distribution of the problem items are thus concentrated in the null category. When the expected cell frequencies are low, the polychoric correlation coefficient may be distorted, unless extremely large samples are used (Muthén, 1989). For these two reasons, estimates of the polychoric correlations are considerably more unstable than the usual PMCCs (Muthén, 1989). In the present study, it was decided to fit the cross-informant model to both polychoric correlations and PMCCs. Prelis-2.12a (Jöreskog & Sörbom, 1993a) was used to calculate both measures of association.

Fit function. The second problem for the analysis of the data reported here concerns the fit function. The choice of fit function is guided by the distribution of the items, where the normal distribution theory estimators (e.g. Maximum Likelihood) apply to multivariate normally distributed items. The Asymptotic Distribution Free (ADF) (Browne, 1984) estimator applies to all other distributions. Theoretically, the ADF estimator is here the appropriate fit function.

However, for practical data analysis, the usefulness of the ADF test statistic is seriously limited because of its extreme instability (Hu, Bentler, & Kano, 1992). The skewness of the data aggravates this problem (Muthén, 1989). Simulation studies have shown that only when sample size is extremely large and/or the number of degrees of freedom are relatively small, does the ADF chi-square statistic work satisfactorily (Muthén & Kaplan, 1985, 1992). The large measurement model (due to the number of items contained in both the CBCL and TRF) evaluated in this study (and consequent large number of degrees of freedom) prohibits the use of the ADF fit function (see Muthén, 1989). It has been suggested that more than 10,000 children for a single sample would be needed to use ADF for the analysis of the cross-informant model (Dedrick et al., 1997).

In the present study the ML fit function was applied to the PMCCs, for pragmatic rather than theoretical reasons. ML is the most commonly used estimation method in factor analysis (Marsh et al., 1998). The ULS estimation method was used for the analysis of the polychoric correlations (see Rigdon & Ferguson, 1991). Lisrel-8.12a (Jöreskog & Sörbom, 1993b) was used for both ML and ULS. The latter method allows for comparisons of results with the earlier-cited studies on the measurement structure of the CBCL and the TRF (Dedrick et al., 1997; De Groot et al., 1994, 1996; Van den Oord, 1993), as these studies consistently used polychoric correlations as the measure of association and ULS as the estimation method. In this study, we will also be able to compare the results from ML estimation and PMCCs with those from ULS estimation and polychoric correlations.

Model Fit

Conventional rules of fit. It was noted above that there is no optimal measure of association and no appropriately defined fit function for the data studied here. Consequently, the calculated chi-square statistic does not follow the theoretical chi-square distribution and is therefore difficult to interpret. The evaluation of how adequately the model fits the data is thus seriously impeded.

Fortunately, the fit of the model to the data may be assessed by other means than the chi-square. Multiple fit indices are generally used because there is no agreed upon best fit index. In addition to chi-square, the following fit indices are considered in the present study: Root Mean Square Error of Approximation (RMSEA) (Steiger, 1990), Root Mean Square Residual (RMR) (Bollen, 1989), Goodness of Fit Index (GFI) (Jöreskog & Sörbom, 1989; Tanaka & Huba, 1985), and the Comparative Fit Index (CFI) (Bentler, 1990).

A major problem with these fit indices is that the theoretical probability distributions for these fit indices are unknown. Consequently, rules of thumb are used for the range of values that are generally taken to indicate a good fit. This concerns the following ranges: RMSEA (0.03–0.07); RMR (0–.05); GFI (.90–1.00); CFI (.90–1.00).

However, the extent to which the data characteristics reported here influence the values of these fit indices cannot be determined. Fixed cutoff values for adequate fit may not work well with large models, large sample sizes, and categorically skewed variables, resulting in the aforementioned less than optimal measures of association and estimation methods. Whether the above-mentioned rules of thumb apply to the present situation is currently unknown.

Simulation

Inadequate values of the various fit indices may result from violation of the factor analytic requirement of multivariate normality of the variables. Thus, inadequate fit values do not unequivocally indicate that the model is wrong, as the skewed, discrete variables analysed here by no means approximate normality. It was, therefore, decided that additional procedures were needed for evaluating goodness of fit. One way to accomplish this is by means of studying the chi-square and the other fit indices in a simulation study.

In simulation, instead of deciding on the fit of the model on the basis of the theoretical distribution of a fit index (chi-square) or on the basis of a priori cutoff values (RMSEA, RMR, GFI, CFI), which may or may not be applicable to the present data, model fit is evaluated by the *empirical* probability distribution of these fit indices. A simulation study provides distributions for the various fit indices taking the skewed categorical distributions observed in the samples into account. The idea of simulation is to draw samples repeatedly with the distribution characteristics as observed in the sample from a population for which the theorised model holds, but with the introduction of random error through sampling. Subsequently, the theorised model is fitted to each of these simulation samples, in order to obtain an empirical sampling distribution of the fit indices. Actual values of the fit indices as they are found for each of the samples studied here may then be compared with this range of values, which fall under random sampling variations if the model is valid. In these simulated distributions of the fit indices, potentially inadequate fit due to inaccuracy of the theorised model is disentangled from apparent inadequate fit caused by violations of distribution assumptions. Thus, these empirical sampling distributions of the fit indices provide a frame of reference by which the fit of the cross-informant model can be evaluated.

In summary, the simulation study is designed such that (1) the distribution characteristics of the items in the simulation samples are like the items in the sample for which the cross-informant model is evaluated, and (2) the simulation samples are drawn from a population that is consistent with the correlational structure implied by the cross-informant model.

To obtain results that are sufficiently precise (see, e.g., Efron & Tibshirani, 1993), 400 simulation samples were drawn for each sample (countries), measure of association (PMCCs and polychoric correlation), model (the common item and the full cross-informant model), and informant (parent and teacher). The sample size of these simulation samples equals the sample size of the sample for which the cross-informant model is evaluated (e.g. 400 simulation samples with sample size $N = 1213$ for the parent sample of Greece). For each sample, measure of association, model, and informant, an empirical probability distribution is provided for chi-square, RMSEA, RMR, GFI, and CFI, based on 400 values resulting from 400 fits of these 2 models to the simulated data.

For polychoric correlations this simulation procedure is well known and available in the computer program Prelis-2.12a (Jöreskog & Sörbom, 1993a) combined with Lisrel-8.12a (Jöreskog & Sörbom, 1993b). First, Lisrel-8.12a was used to generate the model-implied population polychoric correlation matrix $\Sigma(\hat{\theta})$, for which the cross-informant model holds. Second, ordinal data were simulated from this population with Prelis-2.12a following the distributions of the items in the actual samples (Jöreskog & Sörbom, 1993a, pp. 16–21). Third, Lisrel-8.12a was used to fit the cross-informant model with ULS to the polychoric correlations estimated for each of these simulated samples.

For PMCCs the simulation procedure is based on an algorithm originally proposed by Boomsma (1983). Since this procedure is relatively unknown, it is presented here briefly (see also Hox & Hartman, 1999b, for an extensive description). The algorithm starts with the model-implied PMCC population correlation matrix $\Sigma(\hat{\theta})$, for which the cross-informant model holds. It is assumed that the observed skewed, discrete variables with model-implied correlation matrix $\Sigma(\hat{\theta})$ are obtained from a specific categorisation of underlying normally distributed variables with a correlation matrix ρ. The key issue in this procedure is to estimate ρ (see next paragraph). Once ρ is known, simulation is straightforward, because procedures to draw simulation samples from a multivariate normal population with a specified covariance matrix are well known and built into computer programs such as EQS 5.6 (Bentler, 1995). After drawing these simulation samples, the standardised normal variables were subsequently categorised according to the observed category proportions of each of the problem items in the sample under consideration (e.g. the parent sample of Greece). The cross-informant model was subsequently fitted to each of the correlation matrices calculated on the basis of these categorised variables in the simulation samples. Thus, except for random sampling variation, the cross-informant model is consistent with these correlation matrices. The resulting empirical distribution of the fit indices provides the range of values obtained under random sampling of skewed categorically measured variables for which the cross-informant model holds.

The estimation of ρ requires further elaboration. In the above described procedure, the key issue is to estimate a correlation matrix ρ on the basis of which normally distributed simulation samples are drawn, which, after categorisation ($\rho_{\text{categorised}}$), show the correlation patterns implied by the cross-informant model $\Sigma(\hat{\theta})$. For each element of $\Sigma(\hat{\theta})$, i.e. the model-implied correlation between two variables, it is assumed that this correlation results from categorising two underlying continuous variables with a bivariate normal distribution and correlation ρ. The cutting points for the categorisation are estimated from the univariate observed distribution of the variables in the sample under consideration. Under this model, $\rho_{\text{categorised}}$ given ρ is calculated using numerical integration of the underlying bivariate normal distribution. ρ is estimated iteratively, starting with an initial estimate that is equal to $\Sigma(\hat{\theta})$. This estimate is iteratively improved until $\rho_{\text{categorised}}$ differs from the model-implies $\Sigma(\hat{\theta})$ less than a specified criterion of 0.001. ρ serves as the population covariance matrix on the basis of which the simulation samples are drawn (see the above paragraph).

ρ was estimated with the computer program Simulcat (Hox, 1998). Second, EQS 5.6 (Bentler, 1995) was used to draw the simulation samples from the population matrix. Third, these data were categorised with EQS 5.6 following the distribution of the items in the actual samples. Fourth, EQS 5.6 was used to fit the cross-informant model with ML to the PMCC correlation matrices of each of these simulated samples. We used both Prelis-2.12a/Lisrel-8.12a (ULS/polychoric correlations) and EQS 5.6 (ML/PMCCs) to make optimal use of simulation features available in each of these programs.

Comparison with other models. As a third way of deciding on the overall fit of the model, the values of the fit indices are judged comparatively for a series of models.

Five models were considered in which fewer problem dimensions than the cross-informant model are hypothesised.

Model 1, the most restrictive model fitted to the data, is the independence model. This model hypothesises that all problem items in the model are uncorrelated, indicating that no common factors underlie the items. The goodness of fit (or rather the "badness" of fit) provides a measure of the information in the data to be explained by better models, i.e. the lower the fit, the more covariation present in the data. The independence model has the lowest possible fit as compared to models that *do* assume common factors. It can thus be considered as a baseline for evaluating the fit of other models.

Model 2 is a single factor model. This model tests for the possibility that one undifferentiated latent dimension underlies the items.

Model 3 is the eight-factor cross-informant model with uncorrelated factors. This model best represents the cross-informant model as originally derived, since an orthogonal rotation method was used (Varimax) (Achenbach, 1991a, b).

Both the fourth and the fifth models are based on the distinction between internalising and externalising problem behaviour. This has been regarded as a meaningful distinction in child psychopathology (Achenbach & Edelbrock, 1978; Cantwell, 1996; Rutter et al., 1969; Verhulst & Van der Ende, 1992). The fourth and the fifth models are in keeping with Achenbach's grouping of syndromes into internalising and externalising problem behaviour. Items from withdrawn, somatisation, and anxiety/depression load with the internalising factor and items from aggression and delinquency load with the externalising factor. The remaining problem items from social, thought, and attention problems do not pertain to the internalising/externalising distinction (Achenbach, 1991a, b).

Model 4 is a two-factor model in which the remaining problem items from social, thought, and attention problems scales are hypothesised to load with both factors. A study by Song, Singh, and Singer (1994) on the measurement structure of the Youth Self Report (YSR) (Achenbach, 1991c) provided support for this model. In that study, superior model fit was found when social, thought, and attention problems cross-loaded on both the internalising and the externalising problem dimensions. Following Song et al., the two factors were allowed to correlate here. This is consistent with findings that children sometimes show both internalising and externalising behaviour (Angold et al., 1999; Angold & Costello, 1993; Biederman, Faraone, Mick, & Lelon, 1995; Kovacs & Pollock, 1995; Loeber & Keenan, 1994; Loeber, Russo, Stouthamer-Loeber, & Lahey, 1994; McConaughy & Skiba, 1993; Pliszka, 1992; Zoccolillo, 1992).

Model 5 is a five-factor model which specifies social, thought, and attention problems as separate factors, in addition to the internalising and externalising factors. Again, the factors were allowed to correlate.

Model 6 is the eight-factor cross-informant model (see Fig. 1). The eight factors were allowed to correlate. The improvement in fit may be assessed for the cross-informant model over and above the aforementioned models.

Finally, the least restricted model in this series is the unrestricted model (Model 7) (Jöreskog, 1979b). Except for the minimum number of restrictions required for model identification (Jöreskog, 1979b), no specific pattern is specified for the problem items loading with the underlying syndrome dimensions, i.e. all but eight items load on all eight latent variables. The unrestricted model is statistically equivalent to an exploratory factor analysis and gives identical goodness of fit for the data. This model essentially assesses whether the number of factors is appropriate to describe the data adequately, regardless of the pattern of high and low factor loadings (the substantive meaning of the factors). The fit of the unrestricted model indicates the best possible fit for an eight-factor model. The comparison with the cross-informant model provides information on the extent to which fit deteriorates as a consequence of the specific measurement structure of the cross-informant model. A large difference in model fit casts doubt on the hypothesised relationships between the problem items and the underlying syndrome dimensions.

All models were fitted to the data using Lisrel-8.12a (Jöreskog & Sörbom, 1993b).

Results

Aptness of the Cross-informant Model Using Conventional Rules of Fit

Overall model fit of the cross-informant model (Model 6) is presented in Tables 2a and 2b for parent and teacher ratings, respectively. Fit indices are provided for two methods, the PMCCs analysed with ML, and the polychoric correlations analysed with ULS. Two models were evaluated: first, a restricted cross-informant model based on those problem items common to the CBCL and the TRF, and second the full cross-informant model (Achenbach, 1991a, b).

Parent data. For the parent data, the ML estimation method (applied to PMCCs, Table 2a) gave high model chi-square values. Two other fit indices (RMSEA and RMR) provide acceptable to nearly acceptable fit using conventional cutoff scores. The remaining two fit indices, GFI and CFI, are well below the range of values considered acceptable. Similar results are observed for the common item and the full cross-informant models.

Using the second method, ULS estimation (applied to the polychoric correlations, Table 2a), the chi-square values are very high. The remaining four fit indices show the opposite pattern of results compared with the ML method of the PMCCs. The RMSEA and RMR are inadequate, while GFI and CFI are almost acceptable. There is no difference between the common item and the full cross-informant model.

No solution could be found for the Norwegian data using the ULS method nor for the Israeli data for the common item model.

It should be noted that the above pattern of results is consistent across the different countries. No clear-cut differences emerged between population samples and clinical samples.

Teacher data. The fit indices for the teacher data, compared with parental data, are somewhat poorer (Table 2b). Both estimation methods gave high chi-square values. The two fit indices, RMSEA and RMR, approach an acceptable fit for the ML analysis of PMCCs, but again suggest the opposite conclusion for the polychoric correlations analysed with ULS, namely, a poor fit. In contrast, the other two fit indices, CFI and GFI, provide inadequate fit for ML but approximate acceptable fit for the polychoric correlations analysed with ULS. The pattern of results is consistent across the different countries for both referred and nonreferred samples.

As can be seen from Table 2b, the fit indices for the common item model and the full cross-informant model are similar, hence no differentiation between these two models can be made based on these results.

No solution could be found for the full cross-informant model for the ULS estimation method for the Greek, United States, and Israeli teacher data, nor for the common item model for the latter sample.

Table 3 lists the results from previous CFA studies of the CBCL (Dedrick et al., 1997; De Groot et al., 1994; Van den Oord, 1993) and TRF (De Groot et al., 1996). Comparison of these studies with the present study is limited to the full cross-informant model and to the ULS estimation method applied to polychoric correlations. The chi-square was reported in two studies, and the GFI and RMR in four studies. Previous results are very similar to those reported here: RMRs tend to be high, indicating inadequate fit, while GFIs approach acceptable fit. Model fit for the teacher data seemed somewhat poorer than model fit for the parent data.

Using conventional rules of fit, the two methods of analysis produced somewhat conflicting results (see Tables 2a, 2b, and 3). Clearly, more detailed analyses are required to evaluate whether the measurement structure of the CBCL and TRF is a good approximation of the covariance patterns in the data.

Aptness of the Cross-informant Model: A Simulation Study

ML/PMCCs. Separate probability distributions were derived for each sample (countries), informant

Table 2a
Model Fit Indices for Cross-informant Measurement Structure of the CBCL (Model 6)

Model	Common items		Full model	
df	2816		3451	
Method	PMCC/ML	Polych/ULS	PMCC/ML	Polych/ULS
Greece ($N = 1213$)				
χ^2	9518	42,316	11,163	52,235
RMSEA	.044	.11	.043	.11
RMR	.060	.11	.058	.11
GFI	.81	.87	.80	.87
CFI	.68	.86	.67	.86
Portugal ($N = 1375$)				
χ^2	10,402	38,724	12,344	32,153
RMSEA	.044	.10	.043	.078
RMR	.056	.10	.058	.080
GFI	.81	.88	.80	.93
CFI	.72	.87	.71	.93
Turkey ($N = 1564$)				
χ^2	11,821	33,507	14,278	42,907
RMSEA	.045	.084	.045	.086
RMR	.056	.084	.060	.087
GFI	.81	.92	.80	.92
CFI	.67	.92	.66	.92
Norway ($N = 1162$)				
χ^2	9668		11,860	
RMSEA	.046		.046	
RMR	.054	$\Sigma(\hat{\theta})$: npd	.055	$\Sigma(\hat{\theta})$: npd
GFI	.81		.79	
CFI	.65		.62	
Netherlands ($N = 1753$)				
χ^2	16,435	46,872	19,204	55,440
RMSEA	.053	.094	.051	.093
RMR	.071	.094	.070	.093
GFI	.76	.89	.75	.89
CFI	.70	.88	.69	.88
Israel ($N = 2246$)				
χ^2	19,727	49,620	22,860	59,758
RMSEA	.052	.086	.050	.085
RMR	.069	.086	.068	.085
GFI	.78	.91	.76	.91
CFI	.67	.89	.66	.89
Israel ($N = 1340$)				
χ^2	10,756		13,017	93,696
RMSEA	.046		.045	.14
RMR	.054	**	.054	.14
GFI	.81		.79	.81
CFI	.63		.60	.79
United States ($N = 2573$)				
χ^2	15,703	44,605	18,587	53,805
RMSEA	.042	.076	.041	.075
RMR	.048	.076	.048	.076
GFI	.84	.94	.83	.94
CFI	.75	.93	.73	.93

PMCC/ML: Maximum Likelihood estimation method applied to product moment correlation coefficients; polych/ULS: Unweighted Least Squares estimation method applied to polychoric correlations; χ^2 is rounded to the nearest integer; Greek, Portuguese, Turkish, Norwegian, Israeli ($N = 1340$) and United States samples are population based; Dutch and Israeli ($N = 2246$) samples are clinically referred samples.

$\Sigma(\hat{\theta})$: npd: The estimated model correlation matrix was not positive definite (see Wothke, 1993).

**: The solution did not converge for this model.

(parent and teacher), and model (common symptom model and full cross-informant model). Each probability distribution was based on 400 simulation samples. These simulated probability distributions encompass the range of values which indicate adequate fit, against which the validity of the cross-informant model can be assessed.

Table 2b
Model Fit Indices for Cross-informant Measurement Structure of the TRF (Model 6)

Model	Common item model		Full model	
df	2816		4911	
Method	PMCC/ML	Polych/ULS	PMCC/ML	Polych/ULS
Greece ($N = 1179$)				
χ^2	16,846	115,776	26,449	
RMSEA	.065	.18	.061	
RMR	.092	.18	.10	$\Sigma(\hat{\theta})$: npd
GFI	.67	.83	.61	
CFI	.65	.82	.65	
Portugal ($N = 1377$)				
χ^2	17,206	58,083	29,071	122,846
RMSEA	.061	.13	.060	.13
RMR	.086	.13	.096	.13
GFI	.70	.91	.64	.90
CFI	.69	.91	.66	.90
Turkey ($N = 1608$)				
χ^2	18,543	70,985	31,293	133,263
RMSEA	.059	.12	.058	.13
RMR	.083	.12	.092	.13
GFI	.71	.90	.64	.89
CFI	.67	.90	.65	.89
Netherlands ($N = 1418$)				
χ^2	17,691	63,424	27,482	100,306
RMSEA	.061	.12	.057	.12
RMR	.087	.12	.089	.12
GFI	.70	.85	.65	.87
CFI	.67	.83	.67	.86
Israel ($N = 954$)				
χ^2	14,747		22,667	
RMSEA	.067		.062	
RMR	.091	$\Sigma(\hat{\theta})$: npd	.095	$\Sigma(\hat{\theta})$: npd
GFI	.65		.58	
CFI	.70		.68	
United States ($N = 2573$)				
χ^2	29,684	76,909	46,349	
RMSEA	.064	.11	.060	
RMR	.090	.10	.092	$\Sigma(\hat{\theta})$: npd
GFI	.68	.95	.62	
CFI	.73	.95	.73	

See Table 2a for abbreviations. Greek, Portuguese, Turkish, and United States samples are population based; the Dutch sample is clinically referred.

Table 3
Model Fit Indices for Full Cross-informant Model in Previous Studies (Model 6)

	CBCL Van den Oord et al.[a]		CBCL De Groot et al.	CBCL Dedrick et al.	TRF De Groot et al.
	($N = 2148$)[b]	($N = 1387$)	($N = 2335$)	($N = 631$)	($N = 1221$)
χ^2	not reported		100,580	17,018	not reported
df	2458	2458	3451	3451	4911
GFI	.96	.89	.89	.91	.85
RMR	.082	.098	.096	.086	.13

Method is Unweighted Least Squares applied to polychoric correlations; χ^2 is rounded to the nearest integer.
[a] A number of items were removed because of low symptom endorsement. Thus the comparison here is with a somewhat reduced cross-informant model.
[b] This sample is an adoption sample; all other samples are clinically referred samples.

Tables 4a and 4b provide the simulated intervals together with the model fit of the cross-informant model for parents and teachers, respectively. Both the common-symptom and the full cross-informant models were evaluated. All indices of model fit, irrespective of model, informant, or country fall outside the null-distribution's range of values indicating adequate fit. This finding unequivocally indicates that the measurement structure

Table 4a
Model Fit Indices and 99% Null Hypothesis Intervals for Cross-informant Measurement Structure of CBCL (Model 6)

Model	Common item model		Full model	
df	2816		3451	
	Model fit	99% interval	Model fit	99% interval
Greece ($N = 1213$)				
χ^2	9518	3514–4237	11,163	4529–5716
RMSEA	.044	.014–.020	.043	.016–.023
RMR	.060	.027–.031	.058	.028–.032
GFI	.81	.92–.93	.80	.90–.92
CFI	.68	.92–.96	.67	.88–.93
Portugal ($N = 1375$)				
χ^2	10,402	3769–4595	12,344	4894–5993
RMSEA	.044	.016–.021	.043	.017–.023
RMR	.056	.026–.029	.058	.027–.032
GFI	.81	.92–.93	.80	.91–.92
CFI	.72	.92–.96	.71	.90–.94
Turkey ($N = 1564$)				
χ^2	11,821	4065–5050	14,278	5498–6911
RMSEA	.045	.017–.023	.045	.019–.025
RMR	.056	.026–.029	.060	.028–.033
GFI	.81	.92–.94	.80	.91–.92
CFI	.67	.89–.93	.66	.85–.91
Norway ($N = 1162$)				
χ^2	9668	4656–6161	11,860	5941–7650
RMSEA	.046	.024–.032	.046	.025–.032
RMR	.054	.032–.038	.055	.034–.040
GFI	.81	.88–.91	.79	.87–.89
CFI	.65	.81–.87	.62	.77–.85
Netherlands ($N = 1753$)				
χ^2	16,435	2923–3336	19,204	3789–4291
RMSEA	.053	.005–.010	.051	.007–.012
RMR	.071	.020–.022	.070	.020–.023
GFI	.76	.95–.96	.75	.95–.95
CFI	.70	.98–1.00	.69	.98–.99
Israel ($N = 1340$)				
χ^2	10,756	4358–5490	13,017	5667–7038
RMSEA	.046	.020–.027	.045	.022–.028
RMR	.054	.029–.033	.054	.030–.034
GFI	.81	.90–.92	.79	.89–.91
CFI	.63	.83–.89	.60	.80–.87
Israel ($N = 2246$)				
χ^2	19,727	3013–3505	22,860	3807–4417
RMSEA	.052	.006–.010	.050	.007–.011
RMR	.069	.018–.020	.068	.019–.021
GFI	.78	.96–.97	.76	.96–.96
CFI	.67	.98–.99	.66	.98–.99
United States ($N = 2573$)				
χ^2	15,703	3398–3949	18,587	4317–5029
RMSEA	.042	.009–.013	.041	.010–.013
RMR	.048	.018–.020	.048	.018–.020
GFI	.84	.96–.97	.83	.96–.96
CFI	.75	.97–.99	.73	.96–.98

Method is Maximum Likelihood applied to product moment correlation coefficients; number of simulation samples for each model and each country is 400. χ^2 is rounded to the nearest integer.
Greek, Portuguese, Turkish, Norwegian, Israeli ($N = 1340$) and United States samples are population based; Dutch and Israeli ($N = 2246$) samples are clinically referred samples.

of the cross-informant model does not adequately describe the covariance patterns in the current data, above and beyond the lack of fit engendered by the distribution properties of the items.

In Tables 4a and 4b it can be seen that model fit for teachers is somewhat poorer than that for parents. To illustrate this, in Tables 4a and 4b chi-square for the common symptom model is 9518 (interval 3514–4237) for

Table 4b
Model Fit Indices and 99% Null Hypothesis Intervals for Cross-informant Measurement Structure of TRF (Model 6)

Model	Common item model		Full model	
df	2816		3451	
	Model fit	99% interval	Model fit	99% interval
Greece ($N = 1179$)				
χ^2	16,846	5037–6991	29,449	8228–10,313
RMSEA	.065	.026–.035	.061	.024–.031
RMR	.092	.031–.038	.10	.030–.036
GFI	.67	.87–.90	.61	.85–.88
CFI	.65	.86–.92	.65	.88–.92
Portugal ($N = 1377$)				
χ^2	17,206	5057–6102	29,071	8211–9692
RMSEA	.061	.024–.029	.060	.022–.027
RMR	.086	.028–.033	.096	.027–.031
GFI	.70	.90–.92	.64	.88–.90
CFI	.69	.90–.94	.66	.91–.93
Turkey ($N = 1608$)				
χ^2	18,543	5070–6230	31,293	7858–9316
RMSEA	.061	.022–.027	.058	.019–.024
RMR	.086	.026–.030	.092	.025–.028
GFI	.70	.91–.93	.64	.90–.91
CFI	.69	.91–.94	.65	.92–.94
Netherlands ($N = 1418$)				
χ^2	17,691	3065–3588	27,482	5448–6076
RMSEA	.061	.008–.014	.057	.009–.013
RMR	.087	.022–.025	.089	.022–.025
GFI	.70	.94–.95	.65	.92–.93
CFI	.67	.97–.99	.67	.98–.99
Israel ($N = 954$)				
χ^2	14,747	3973–4847	22,667	7004–8090
RMSEA	.067	.021–.028	.062	.021–.026
RMR	.091	.027–.032	.095	.028–.032
GFI	.65	.87–.91	.58	.86–.88
CFI	.70	.93–.96	.68	.92–.95
United States ($N = 2357$)				
χ^2	29,684	5637–6812	46,349	9375–10,945
RMSEA	.064	.021–.025	.060	.020–.023
RMR	.090	.021–.025	.092	.021–.024
GFI	.68	.92–.94	.62	.90–.92
CFI	.73	.95–.96	.73	.95–.96

Method is Maximum Likelihood applied to product moment correlation coefficients; number of simulation samples for each model and each country is 400; χ^2 is rounded to the nearest integer.

Greek, Portuguese, Turkish, and United States samples are population based; the Dutch sample is a clinically referred sample; the Israeli sample is a mixed sample.

Greek parents and 16,846 (interval 5037–6991) for Greek teachers. The model for parents differs less than that of the teachers. A similar conclusion holds for the other fit indices. Thus, both models fit poorly but the teacher model diverges somewhat more than that of the parent model from the expected values under the cross-informant model.

Tables 4a and 4b illustrate that model fit is poorer for clinically referred samples than for population-based samples. As an illustration, chi-square for the common symptom CBCL model is 11,821 (interval 4065–5050) for the Turkish population sample and 16,435 (interval 2923–3336) for the Dutch clinical sample. Model fit differs less from the range of values indicating adequate fit for the population sample than for the clinical sample. A similar conclusion holds for the other fit indices. Thus, both are poor fits, but the clinical sample diverges more than the population sample from the expected values under the cross-informant model.

ULS/Polychoric correlations. It was not possible to derive probability distributions of the fit indices for the polychoric correlations evaluated with ULS. For the vast majority of samples no solution could be found when the cross-informant model was fitted to the simulated data. The teacher data were in this respect even more problematic than the parent data. Since symptom endorsement was lower for the teachers, this suggests that the estimation problems are due to the skewed data.

In the first stage of the simulation procedure, standard normally distributed variables were generated and transformed such that, except for sampling variation, their covariance structure was in agreement with the cross-informant measurement structure. For the purposes of locating the cause of the estimation problems, the cross-

Table 5
Comparative Factor Models

Model	Properties
Model 1	
Independence model	Assumes no covariation among the problem items and hence no underlying problem dimensions. Indicates lowest level of fit for these problem items.
Model 2	
Single-factor model	Assumes a single undifferentiated psychopathology factor underlying the problem items, as reported by the informant.
Model 3	
Orthogonal eight-factor model	Assumes uncorrelated factors but is otherwise identical to the cross-informant model.
Model 4	
Two-factor model	Assumes no differentiation within internalising and externalising problem dimensions, i.e. withdrawn, somatic complaints, and anxious/depressed are represented as a single factor and delinquency and aggression are represented as a second factor. Social problems, thought problems, and attention problems load on both the internalising and externalising factor. The two factors are allowed to correlate.
Model 5	
Five-factor model	Identical to two-factor model regarding the internalising and externalising distinction. In contrast, social problems, thought problems, and attention problems do not load on the internalising-externalising factors but are represented as separate factors. The five factors are allowed to correlate.
Model 6	
Cross-informant model	Assumes eight correlated problem dimensions (see Fig. 1).
Model 7	
Unrestricted model	Assumes eight factors underlying the problem items but leaves unspecified which symptoms load with which factors (Jöreskog, 1978b). The eight factors are allowed to correlate. Indicates upper level of fit for an eight-factor model.

informant model was fitted to the PMCCs calculated for these normally distributed data. No estimation problems occurred in this phase.

In the second stage, these simulated data were trichotomised according to the distribution of each of the problem items in the actual sample. When the cross-informant model was fitted to the polychoric correlations estimated from these categorised data, the estimation problems emerged. Removal of the most skewed symptoms resulted in convergence of the model fitting process in most samples.

These results again suggest that accurate estimation of the population polychoric correlations may not be possible for extremely skewed categorically measured data (see Muthén, 1989; Muthén et al., 1993). A small change in the number of children in the 1 and 2 categories of the distribution may result in a large change in the estimated values of the polychoric correlation with other variables. The results suggest that sampling variability caused the polychoric correlations to deviate from the cross-informant measurement structure to the extent that no solutions could be obtained.

Aptness of Cross-informant Model: Comparison with Other Models

The cross-informant model was compared with a number of alternative models. Table 5 provides a description of these models. Fit indices for these models are presented in Tables 6a and 6b, for parent and teacher data respectively. Results are provided for both PMCCs analysed with ML and polychoric correlations analysed with ULS. Tables 6a and 6b are based on the problem items of the full cross-informant model. (Similar tables based on the problem items of the common symptom model may be obtained from the first author.)

Results were similar for parent and teacher data, for the two methods of analysis, and for the common symptom model as well as the full cross-informant model.

The independence model (Model 1) shows extremely poor fit in all instances, indicating that there is considerable covariance among the problem items, which needs to be explained.

The single factor model (Model 2) shows a large improvement in fit as compared with the independence model. This result indicates that a considerable part of the covariation is explained by one undifferentiated factor.

For the orthogonal cross-informant model (Model 3), large residuals (RMR) were found. These residuals approach those of the independence model, which indicates the lowest possible fit for these items. Again, this finding is indicative of substantial covariance underlying the problem items. The poor fit of the orthogonal cross-informant model becomes worse for the polychoric correlations analysed with ULS. In a number of instances no solution could be found for this method.

In comparison to the single-factor model, the two-factor model (Model 4) shows some improvement in fit. This suggests some support for the distinction between internalising and externalising problem behaviour, particularly for teachers.

The goodness of fit of the five-factor model (Model 5)

Table 6a
Model Fit for Comparative Factor Models for Parent Ratings

df	Independence Model 1 3570	1-factor Model 2 3485	Orthogonal Model 3 3479	2-factor Model 4 3461	5-factor Model 5 3465	C.I. Model 6 3451	Unrestricted Model 7 2918
Greece ($N = 1213$)							
ML/PMCCS							
χ^2	27,084	13,961	14,257	11,980	11,856	11,163	6695
RMSEA	.074	.050	.051	.045	.045	.043	.033
RMR	.15	.057	.13	.052	.057	.058	.029
GFI	.34	.72	.74	.79	.79	.80	.88
CFI	.00	.55	.54	.64	.64	.67	.84
ULS/Polych							
χ^2	353,240	65,943	263,561	54,254	54,656	52,235	
RMSEA	.28	.12	.25	.11	.11	.11	
RMR	.28	.12	.24	.11	.11	.11	**
GFI	.13	.84	.35	.87	.86	.87	
CFI	.00	.82	.26	.85	.85	.86	
Portugal ($N = 1375$)							
ML/PMCCS							
χ^2	34,462	17,092	16,268	14,237	13,774	12,344	6779
RMSEA	.079	.053	.052	.048	.047	.043	.031
RMR	.17	.064	.14	.056	.057	.058	.026
GFI	.28	.67	.73	.75	.76	.80	.89
CFI	.00	.56	.59	.65	.67	.71	.88
ULS/Polych							
χ^2	435,709	49,784		35,734	35,660	32,153	
RMSEA	.30	.10		.082	.082	.078	
RMR	.29	.10	$\Sigma(\hat{\theta})$: npd	.084	.084	.080	**
GFI	.12	.90		.93	.93	.93	
CFI	.00	.89		.93	.93	.93	
Turkey ($N = 1564$)							
ML/PMCCS							
χ^2	35,287	19,375	18,402	16,419	16,429	14,278	7694
RMSEA	.075	.054	.052	.049	.049	.045	.032
RMR	.15	.065	.13	.058	.059	.060	.027
GFI	.34	.68	.73	.75	.75	.80	.89
CFI	.00	.50	.53	.59	.59	.66	.85
ULS/Polych							
χ^2	488,334	62,748		48,170		42,907	
RMSEA	.29	.10		.091		.086	
RMR	.29	.10	$\Sigma(\hat{\theta})$: npd	.092	$\Sigma(\hat{\theta})$: npd	.087	**
GFI	.12	.89		.91		.92	
CFI	.00	.88		.91		.92	
Norway ($N = 1162$)							
ML/PMCCS							
χ^2	25,426	14,753	14,229	12,776	12,643	11,860	7562
RMSEA	.073	.053	.052	.048	.048	.046	.037
RMR	.14	.060	.11	.054	.056	.055	.032
GFI	.38	.71	.74	.77	.77	.79	.86
CFI	.00	.48	.51	.57	.58	.62	.79
ULS/Polych							
χ^2	465,443	153,736		137,161			
RMSEA	.33	.19		.18			
RMR	.33	.19	$\Sigma(\hat{\theta})$: npd	.18	$\Sigma(\hat{\theta})$: npd	$\Sigma(\hat{\theta})$: npd	**
GFI	.096	.70		.73			
CFI	.00	.67		.71			
Netherlands ($N = 1753$)							
ML/PMCCS							
χ^2	53,620	30,664	23,251	23,382	21,925	19,204	9796
RMSEA	.089	.067	.057	.057	.055	.051	.037
RMR	.18	.086	.14	.070	.073	.070	.027
GFI	.26	.53	.70	.69	.71	.75	.87
CFI	.00	.46	.60	.60	.63	.69	.86
ULS/Polych							
χ^2	434,497	97,466	271,199	61,022	61,999	55,440	
RMSEA	.26	.12	.21	.097	.098	.093	
RMR	.26	.12	.21	.098	.098	.093	**

Table 6a (*cont.*)

	Independence	1-factor	Orthogonal	2-factor	5-factor	C.I.	Unrestricted
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
df	3570	3485	3479	3461	3465	3451	2918
GFI	.15	.81	.47	.88	.87	.89	
CFI	.00	.78	.38	.87	.86	.88	
Israel ($N = 2246$)							
ML/PMCCS							
χ^2	60,081	33,188	28,417	27,211	26,368	22,860	11,579
RMSEA	.084	.062	.057	.055	.054	.050	.036
RMR	.17	.073	.14	.065	.070	.068	.027
GFI	.29	.59	.71	.72	.72	.76	.88
CFI	.00	.47	.56	.58	.59	.66	.85
ULS/Polych							
χ^2	536,552	93,943	368,364	65,987		59,758	
RMSEA	.26	.11	.22	.090		.085	
RMR	.26	.11	.21	.090	**	.085	**
GFI	.15	.85	.42	.90		.91	
CFI	.00	.83	.32	.88		.89	
Israel ($N = 1340$)							
ML/PMCCS							
χ^2	27,479	15,198	16,352	13,817	13,697	13,017	8064
RMSEA	.071	.050	.053	.047	.047	.045	.036
RMR	.14	.056	.12	.052	.054	.054	.032
GFI	.39	.74	.73	.78	.78	.79	.87
CFI	.00	.51	.46	.57	.57	.60	.78
ULS/Polych							
χ^2	434,500	105,685	336,651	96,018	97,108	93,696	
RMSEA	.30	.15	.27	.14	.14	.14	
RMR	.30	.15	.26	.14	.14	.14	**
GFI	.12	.78	.31	.80	.80	.81	
CFI	.00	.76	.23	.79	.78	.79	
United States ($N = 2573$)							
ML/PMCCS							
χ^2	60,050	26,012	25,986	22,212	21,260	18,587	10,310
RMSEA	.078	.050	.050	.046	.045	.041	.031
RMR	.17	.054	.14	.049	.049	.048	.025
GFI	.28	.73	.76	.79	.80	.83	.91
CFI	.00	.60	.60	.67	.68	.73	.87
ULS/Polych							
χ^2	754,539	78,516	539,839	61,930	60,876	53,805	
RMSEA	.29	.091	.24	.081	.080	.075	
RMR	.28	.091	.24	.081	.080	.076	**
GFI	.13	.91	.38	.93	.93	.94	
CFI	.00	.90	.29	.92	.93	.93	

Chi-squares are rounded to the nearest integer.

$\Sigma(\hat{\theta})$: npd: The estimated model correlation matrix is not positive definite (see Wothke, 1993).

**: The solution did not converge for this model.

Models are based on the 85 problem items of the full CBCL cross-informant model.

Greek, Portuguese, Turkish, Norwegian, Israeli ($N = 1340$), and US samples are population based; Dutch and Israeli ($N = 2246$) samples are clinically referred samples.

is very similar to that of the two-factor model. No change in fit is observed whether social problems, thought problems, and attention problems are represented as separate factors or whether they are specified as loading on both internalising and externalising problem dimensions. For a number of samples no solution could be found for the polychoric correlations analysed by ULS.

The cross-informant model (Model 6) shows a minor improvement compared with the two- and five-factor model. This shows that the differentiation of a crude internalising problem dimension into more specific types of internalising problem behaviour, i.e. withdrawn, somatisation, and anxiety/depression, is not strongly supported by the data. A similar conclusion holds for the distinction of externalising behaviour into aggression and delinquency.

The unrestricted model (Model 7) shows considerable improvement in fit compared with the cross-informant model. The unrestricted model evaluates whether eight factors are in principle an adequate number to explain the covariance patterns of the data without imposing additional restrictions as to which problem items load with which factors. The improvement in fit for the unrestricted model compared with the cross-informant model suggests that there is misspecification in the measurement structure of the CBCL and the TRF. No solutions could be found for this model for the polychoric correlations analysed with ULS.

Table 6b
Model Fit for Comparative Factor Models for Teacher Ratings

	Independence Model 1	1-factor Model 2	Orthogonal Model 3	2-factor Model 4	5-factor Model 5	C.I. Model 6	Unrestricted Model 7
df	5050	4949	4939	4909	4925	4911	4270
Greece ($N = 1179$)							
ML/PMCCS							
χ^2	66,833	39,141	30,237	29,331	28,400	29,449	13,629
RMSEA	.10	.077	.066	.065	.064	.061	.043
RMR	.24	.10	.19	.085	.10	.10	.029
GFI	.15	.36	.57	.52	.58	.61	.81
CFI	.00	.45	.59	.60	.62	.65	.85
ULS/Polych							
χ^2	1,044,338	212,328		159,271			
RMSEA	.42	.19		.16			
RMR	.41	.19	$\Sigma(\hat{\theta})$: npd	.16	$\Sigma(\hat{\theta})$: npd	$\Sigma(\hat{\theta})$: npd	**
GFI	.05	.81		.86			
CFI	.00	.80		.85			
Portugal ($N = 1377$)							
ML/PMCCS							
χ^2	75,909	42,427	33,410	32,037	31,081	29,071	15,811
RMSEA	.10	.074	.065	.063	.062	.060	.044
RMR	.24	.097	.19	.081	.093	.096	.028
GFI	.15	.39	.58	.58	.61	.64	.80
CFI	.00	.47	.60	.62	.63	.66	.84
ULS/Polych							
χ^2	1,171,345	171,753		114,792		122,846	
RMSEA	.41	.16		.13		.13	
RMR	.41	.16	$\Sigma(\hat{\theta})$: npd	.13	$\Sigma(\hat{\theta})$: npd	.13	**
GFI	.056	.86		.91		.90	
CFI	.00	.86		.91		.90	
Turkey ($N = 1608$)							
ML/PMCCS							
χ^2	79,797	45,761	37,140	33,279	33,288	31,293	14,925
RMSEA	.096	.072	.064	.060	.060	.058	.039
RMR	.22	.097	.18	.079	.092	.092	.026
GFI	.16	.39	.59	.57	.61	.64	.83
CFI	.00	.45	.57	.62	.62	.65	.86
ULS/Polych							
χ^2	1,176,981	185,621		117,278		133,263	
RMSEA	.38	.15		.12		.13	
RMR	.38	.15	$\Sigma(\hat{\theta})$: npd	.12	$\Sigma(\hat{\theta})$: npd	.13	**
GFI	.065	.85		.91		.89	
CFI	.00	.85		.90		.89	
Netherlands ($N = 1418$)							
ML/PMCCS							
χ^2	72,897	40,671	31,172	33,269	30,338	27,482	13,836
RMSEA	.097	.071	.061	.064	.060	.057	.040
RMR	.22	.094	.17	.088	.092	.089	.028
GFI	.15	.43	.60	.55	.61	.65	.82
CFI	.00	.47	.61	.58	.63	.67	.86
ULS/Polych							
χ^2	693,853	138,836	410,758	104,430	106,958	100,306	
RMSEA	.31	.14	.24	.12	.12	.12	
RMR	.31	.14	.24	.12	.12	.12	**
GFI	.093	.82	.46	.86	.86	.87	
CFI	.00	.81	.41	.86	.85	.86	
Israel ($N = 954$)							
ML/PMCCS							
χ^2	60,809	32,578	27,789	25,435	24,453	22,667	11,804
RMSEA	.11	.077	.070	.066	.065	.062	.043
RMR	.28	.098	.24	.086	.095	.095	.028
GFI	.11	.35	.53	.50	.55	.58	.79
CFI	.00	.50	.59	.63	.65	.68	.86
ULS/Polych							
χ^2	867,315	100,109		73,292			
RMSEA	.42	.14		.12			
RMR	.42	.14	$\Sigma(\hat{\theta})$: npd	.12	$\Sigma(\hat{\theta})$: npd	$\Sigma(\hat{\theta})$: npd	**

Table 6b (*cont*.)

df	Independence Model 1 5050	1-factor Model 2 4949	Orthogonal Model 3 4939	2-factor Model 4 4909	5-factor Model 5 4925	C.I. Model 6 4911	Unrestricted Model 7 4270
GFI	.053	.89		.92			
CFI	.00	.89		.92			
United States ($N = 2357$)							
ML/PMCCS							
χ^2	156,182	74,900	56,165	60,748	52,685	46,349	22,120
RMSEA	.11	.077	.066	.069	.064	.060	.042
RMR	.30	.096	.24	.086	.090	.092	.024
GFI	.099	.37	.57	.46	.58	.62	.82
CFI	.00	.54	.66	.63	.68	.73	.88
ULS/Polych							
χ^2	2,634,919	215,602		135,384			
RMSEA	.47	.13		.11			
RMR	.47	.13	$\Sigma(\hat{\theta})$: npd	.11	$\Sigma(\hat{\theta})$: npd	$\Sigma(\hat{\theta})$: npd	**
GFI	.043	.92		.95			
CFI	.00	.92		.95			

Chi squares are rounded to the nearest integer.
$\Sigma(\hat{\theta})$: npd: The estimated model correlation matrix is not positive definite (see Wothke, 1993).
**: The solution has not converged for this model.
Models are based on the 101 problem items of the full TRF cross-informant model.
The Greek, Portuguese, Turkish, Israeli, and US samples are population based; the Dutch sample is a clinically referred sample.

Based on the comparisons between this series of models, we do not find strong support for the differentiation between the eight syndrome dimensions of the CBCL and the TRF.

Discussion

In this paper the internal construct validity of the cross-informant model of the CBCL and the TRF was evaluated using CFA. Using conventional cutoff scores for assessing model fit, it was found that different methods and fit indices provided somewhat conflicting results. For ML, RMSEA and RMR approached adequate fit for the cross-informant model, whereas GFI and CFI indicated inadequate fit. In contrast, for ULS, GFI and CFI suggested almost adequate fit for the cross-informant model, whereas RMSEA and RMR indicated inadequate fit. Since there is no agreed best method for factor analysing the data reported here, these results indicate that reliance on a single method or fit index is unwarranted. In order not to be dependent on conventional rules of fit, which may not be applicable to the present data, empirical probability distributions of the fit indices were derived in a simulation study. It was shown that the fit indices as they were found for the cross-informant model were well outside the range of values indicating adequate fit. Hence, the cross-informant model was unequivocally rejected. However, it could be argued that, given the large model, adequate fit is not a realistic goal (see Marsh et al., 1998). Therefore, in addition to interpretation of goodness of fit in absolute terms, the explanatory value of the cross-informant model was examined as compared to simpler models. The results showed a general dominance of a single factor and a negligible improvement in model fit for the cross-informant model as compared with the internalising and externalising problem dimensions. Thus, these results indicate poor conceptual differentiation and little empirical evidence as to how the cross-informant syndromes are indexed by which items. These results were consistent across countries, informants, and both population and clinical samples[1]. In view of the differences between the present and past reports on the cross-informant model of

[1] One anonymous reviewer suggested that the cross-informant model as formulated (Achenbach, 1991a, b) is too stringent a test of the proposed structure of the CBCL and the TRF. It was proposed that a more appropriate model would be one that allows the specific factors to be mutually correlated. We agree with the argument that there are many reasons why test items may be correlated above and beyond the more substantive factors of interest in the measurement instrument (e.g. difficulty factors, synonyms, etc.). Therefore, we explored the possibility of an adequate model fit for a correlated uniqueness cross-informant model. However, based on the ML/PMCC method and the common item model, model fit did not increase to any great extent for any of the samples when all correlated errors $\geqslant |.20|$ were modelled. The reviewer's second suggestion with regard to correlated errors concerned the comparison with alternative models (e.g. a two-factor model). It was argued quite rightly that unmodelled correlated errors in any model decreases the potential to discriminate between them. To explore this, we used the sample with the largest number of correlated errors for the cross-informant model (Greek teacher sample), and compared the goodness of fit with the two-factor model, for which, similarly, all correlated errors $\geqslant |.20|$ were modelled. Model fit was slightly better for the more parsimonious two-factor correlated uniqueness model (11 correlated errors, $df = 2806$, $\chi^2 = 13{,}731$, RMSEA = .057, RMR = .082, GFI = .73, CFI = .72) as compared to the cross-informant correlated uniqueness model (18 correlated errors, $df = 2798$, $\chi^2 = 14{,}150$, RMSEA = .059, RMR = .085, GFI = .72, CFI = .71). It was not realistic to model the correlated errors for the ULS/polychoric correlations method, because of the extremely large number of correlated errors $\geqslant |.20|$ present, which is most probably due to the erratic behaviour of the polychoric correlation coefficient. The issue of correlated errors could obviously be explored in much more detail. As it stands, this result supports our claim that the cross-informant model does not show adequate construct differentiation.

the CBCL (Achenbach, 1991a, b; Berg et al., 1997; Dedrick et al., 1997; De Groot et al., 1994; Macmann et al., 1992; Van Den Oord, 1993) and the TRF (De Groot et al., 1996), these studies are discussed below.

The original factorial structure of the CBCL and the TRF was developed with PCA (Achenbach, 1991a, b). PCA does not evaluate model fit. Rather, PCA *identifies* possible dimensions that account for covariation among items. The eight syndromes of the CBCL and the TRF were derived on the basis of replication in different samples (Achenbach, 1991a, b). However, the process of identifying, refining, and redefining constructs may proceed slowly and extend across many more subsequent studies, since the conceptual boundaries may only be dimly perceived in the first stages of this research (Comrey & Lee, 1992). A PCA (cf. EFA) may give only a rough idea of the underlying dimensions. A follow-up CFA allows this preliminary model to be refined more precisely. A distinction should therefore be made between the possible identification of problem dimensions and the precision with which these dimensions are conceptualised and measured, as evaluated by CFA. In CFA, the hypothesised cross-informant model is tested for its fit with the observed covariance structure of the problem items. The poor fit reported here suggests little support for the cross-informant syndromes and their differentiation as currently defined.

De Groot et al. (1994, 1996) derived a Dutch model for the CBCL and the TRF. In the first phase of these studies, the emphasis was on the identification of syndrome dimensions. De Groot et al. followed an exploratory approach, subjecting half of the sample to an EFA, to identify a model for the Dutch sample. In the second phase of this research, both the Dutch and the cross-informant (Achenbach, 1991a, b) models were fitted to the remaining half of the Dutch sample by means of CFA. For the CBCL, it was found that the eight-factor solution of the EFA was similar in content to the eight cross-informant syndromes. This result was chosen to represent the Dutch model. In contrast, for the TRF a 12-factor solution was required to identify 8 factors that were similar in content to the cross-informant syndromes. Thus, four additional factors were present in the TRF data, which were not modelled in the eight-factor Dutch model, which was subsequently fitted to the remaining part of the sample. Consistent with this, De Groot et al. found a poorer fit for the Dutch TRF model than for the Dutch CBCL model (see Table 3 here).

It could be argued that the poorer fit indices for the teacher model (De Groot et al., 1996) were caused by a greater violation of distribution assumptions, since teacher ratings are generally more skewed than parent ratings (present study; see also Spiker, Kraemer, Constantine, & Bryant, 1992). However, these differences in skew were incorporated here in the simulation study and, despite this, a slightly poorer fit was found for the TRF compared with the CBCL (see Tables 4a and 4b).

De Groot et al. (1994, 1996) showed that for both the CBCL and the TRF the fit indices of the cross-informant model were nearly identical to those of the Dutch CBCL and TRF models. The authors interpreted this to imply that "the cross-informant syndromes transcend differences in language, culture and mental health systems between Holland and the United States". Identical fit could be interpreted equally well, however, as a relatively arbitrary composition of the items in the scales. For example, in the Dutch CBCL model the items "brags" and "disobedient at school" are part of the delinquent behaviour syndrome; "jealous" is part of the anxious/depressed syndrome; "fights" and "attacks people" are part of the social problems syndrome. In contrast, these five problem items are part of the aggressive behaviour syndrome in the cross-informant model. Similar examples of exchangeable problem items hold for other syndrome dimensions. Thus, on the basis of the covariation patterns in this Dutch sample, it could not be determined whether the Dutch model or the cross-informant model provide a better model for the data, since identical fits were found. Consequently, this result may be interpreted as imprecise measurement of the diagnostic constructs rather than cross-cultural robustness, because of a relative arbitrariness of the problem items in the scales.

Furthermore, a relatively arbitrary construct representation of the CBCL and the TRF is consistent with the present findings and those from the study by Dedrick et al. (1997). Dedrick et al. compared the cross-informant model with three alternative models: the independence model, the single-factor, and the orthogonal eight-factor model. This showed the same pattern of results as was repeatedly found for the samples analysed here. Considerable improvement in fit was found for the single-factor model over both the independence model and the orthogonal eight-factor model. In contrast, improvement in fit for the cross-informant model compared with the single-factor model was small. Taken together, these results indicate that a large single factor dominates the problem items. This factor may in part be due to a halo effect (Epkins, 1994), which is a threat to the constructs measured by the instrument, since they cannot be differentiated adequately from one another. The halo effect may be defined as a rater's failure to discriminate among conceptually distinct and possibly independent aspects of the ratee's performance which, in turn, results in higher correlations among rating dimensions than the true levels of these correlations (Pulakos, Schmitt, & Ostroff, 1986, p. 29). An alternative explanation is that the instrument(s) is measuring one general psychopathology factor (Macmann et al., 1992).

In addition to the single-factor model comparison, the cross-informant model was compared here with two models based on the internalising and externalising dimensions. It was found that model fit slightly improved over the single-factor model, suggesting some but not considerable support for the internalising-externalising distinction. Macmann et al. (1992) showed that for a two-factor model (PCA, biquartimin rotation), the majority of problem items of the CBCL had factor loadings $\geqslant .40$ on both the internalising and externalising dimensions. On the basis of this result, Macmann et al. concluded that the CBCL does not reliably distinguish internalising from externalising problem dimensions. Further differentiation into eight syndrome dimensions was not supported by the results reported in the present study, since improvement in model fit of the cross-informant model compared with the internalising-externalising models was negligible.

A recent study on the convergence between the cross-informant syndromes of the CBCL and clinical diagnoses showed that each CBCL scale predicted a broad range of DSM-III-R diagnoses (Kasius, Ferdinand, van den Berg, & Verhulst, 1997). This finding was attributed to high comorbidity, intrinsic to childhood psychopathology. While recognising the presence of high comorbidity in childhood psychopathology (Angold et al., 1999), the low specificity of the CBCL scales with regard to widely varying DSM diagnoses additionally suggests insufficient construct differentiation in the CBCL. Consistent with this, Lachar (1998) pointed out that the primary evidence of the validity of the CBCL and TRF, as reported in the 1991 manuals (Achenbach, 1991a, b), is that syndrome scales differentiate between clinically referred and normative samples. However, the effectiveness of individual scales with regard to making specific distinctions between different clinical groups was scarcely documented (Lachar, 1998).

The cross-informant model (Achenbach, 1991a, b) was derived following a procedure that effects goodness of fit in CFA. Items that loaded $\geqslant .40$ on the aggressive syndrome and $\geqslant .30$ on a second syndrome were retained only for the second syndrome. This procedure ignores the fact that these items measure two factors rather than one. Further, these items load more on aggression than on the syndrome dimensions to which they were actually assigned in the cross-informant model. Here, the fit of the unrestricted model, which imposes the minimum number of restrictions as to which problem items load with which factors, consistently showed the presence of substantial misspecification and/or cross-loadings in the cross-informant model. This finding may, at least in part, be explained by the orthogonal rotation procedure originally used by Achenbach to derive the cross-informant syndromes. Problem dimensions in child psychopathology are known to be highly correlated (Angold et al., 1999). In the present study, the fit indices for the orthogonal cross-informant model (Model 2) do, in fact, show that this model strains reality. When an orthogonal rotation method in a PCA (cf. EFA) is used, the covariation present in the data comes through in a less conceptually clean factor structure, i.e. incorrect classification of items on factors and/or a large amount of substantial dual or multiple loadings (Cattell & Dickman, 1962; see also Cattell, 1973).

Thus, Van den Oord (1993) allowed a large number of items to (1) have secondary or more cross-loadings on other syndrome dimensions, and (2) load on different syndrome dimensions than originally specified in the cross-informant model. These revisions of the cross-informant model change the conceptual meaning of the syndrome dimensions and the boundaries between them. The resulting improvement in model fit found by Van den Oord further indicates a lack of empirical support for the cross-informant syndromes as currently defined.

A modified model that was based on the best items (47 out of 85) was used in a subsequent study by Van den Oord, Verhulst, and Boomsma (1994). On the basis of both Dutch and French data, Berg et al. (1997) proposed a reduced model using 43 out of 85 problem items. De Groot et al. (1994) also referred to the more robust version of the cross-informant syndromes consisting of the overlapping items between the Dutch measurement structure and the U.S. based cross-informant model. However, the overlap of "robust" CBCL items for these three studies is small, reducing the cross-informant model to four items (aggression) or three items (remaining scales) for each dimension (factor loadings $\geqslant .40$ on the appropriate factor in three studies; Van den Oord, 1993, was used for this purpose, because the factor solution was not reported for the 47-item version that was used in the Van den Oord et al., 1994, study). The small item-overlap across studies suggests a loose anchoring of most problem items in the cross-informant syndrome scales. The construct validity of these reduced factors as measured by the remaining few problem items remains to be determined.

From a conceptual point of view, loose anchoring of items in the scale is consistent with what Kamphaus and Frick (1996) refer to as a lack of coherence of the CBCL and TRF cross-informant syndromes: the item content of the problem dimensions tends to be heterogeneous, leading to problems of interpretation (see also Lachar, 1998). Kamphaus and Frick stress that in order to understand the meaning of an elevated syndrome scale score, it is imperative to view which individual items caused the elevation. However, the sole purpose of summing items into syndrome scores is to yield a more reliable and conceptually more meaningful score than any of the individual item scores.

A second criticism made by Kamphaus and Frick (1996) concerns the conceptual differentiation between the scales of the CBCL and TRF. They argue that the combination of constructs such as anxiety and depression, and hyperactivity and inattention, into single scales hinders differential diagnosis. The inductive approach used for the derivation of the cross-informant syndromes assumes that one may proceed from problem items to adequate syndrome dimensions. However, the adequacy of the dimensions that emerge is a function of the original item pool. In this sense, an inadequate item pool may lead to a lack of conceptual differentiation. The inductive method of questionnaire construction has, moreover, been associated with the following aspects of measurement imprecision: conceptual overlap and imprecise boundaries among the constructs, heterogeneous items that do not necessarily have a clear substantive link to the construct, and substantial method (e.g. halo) rather than construct specific variance (Jackson, 1971). This is consistent with the findings for the cross-informant model discussed here.

Although the CBCL and TRF development was never intended to replace a clinical diagnosis (Achenbach, 1995), there has been a tendency in clinical practice to assume that the syndrome dimensions generated from these instruments are indeed clinical ones. This unfortunate practice should be avoided. Even accepting this point, the meaning of the peaks and the troughs in the CBCL and TRF profiles is obscure, because there is little evidence to support the homogeneity and differentiation of the eight syndrome dimensions. Therefore, it is precarious to interpret differences in scale scores. Furthermore, in research, the power of any study that is aimed at the understanding of childhood psychiatric syndromes depends on the rigour with which the diagnostic groups were defined and selected. Therefore, based on the present results, selection of groups on the basis of high scores on individual CBCL and TRF syndrome scales may be far from optimal because of insufficient measurement precision. That is, scale scores will show too low an association with variables of interest and too high an association with irrelevant variables.

In sum, the present CFA study evaluated the internal validity of the construct representation of the CBCL and the TRF, and consistently showed inadequate empirical support for the cross-informant syndromes.

Acknowledgements—We gratefully acknowledge the comments of Stan Mulaik and Conor Dolan on an earlier draft of this article. For providing the Dutch data, we would like to thank Boudewijn Gunning and Grard Akkerhuis (Academic Medical Centre, Amsterdam), Simone Zijlmans (Municipal Paedological Institute, Amsterdam), Rob Drost (Juliana Hospital for Children, The Hague), Harry Schut (Regional Institute for Mental Welfare [RIAGG] Westhage, The Hague), and Kees de Jonge and Rieta van der Veer (Regional Institute for Mental Welfare [RIAGG] Haagrand, Voorburg/Zoetermeer). We wish to acknowledge the data from the Pittsburgh Youth Study supplied to us by Rolf Loeber. We gratefully acknowledge Novartis for the financial support to the meetings of the European Network for Hyperkinetic Disorder (Eunethydis), which were the basis on which this research project was initiated. Finally, we want to thank three anonymous reviewers and the editor for their valuable comments.

References

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4–18 and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991c). *Manual for the Youth Self-Report and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1995). Empirically based assessment and taxonomy: Applications to clinical research. *Psychological Assessment*, *7*, 261–274.

Achenbach, T. M., & Edelbrock, C. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin*, *85*, 1275–1301.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213–232.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., revised). Washington, DC: Author.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Angold, A., & Costello, E. J. (1993). Depressive comorbidity in children and adolescents: Empirical, theoretical, and methodological issues. *American Journal of Psychiatry*, *150*, 1779–1791.

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry*, *40*, 57–87.

Auerbach, J., Goldstein, E., & Elbadour, S. (1998). *Behavior problems in Bedouin elementary school children*. Manuscript submitted for publication.

Babakus, E., Ferguson, C. E., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, *24*, 222–228.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *88*, 588–606.

Bentler, P. M. (1995). *EQS Structural Equations Program Manual*. Encino, CA: Multivariate Software.

Berg, I., Fombonne, E., McGuire, R., & Verhulst, F. (1997). A cross-cultural comparison of French and Dutch disturbed children using the Child Behaviour Checklist (CBCL). *European Child and Adolescent Psychiatry*, *6*, 7–11.

Biederman, J., Faraone, S., Mick, S., & Lelon, E. (1995). Psychiatric comorbidity among referred juveniles with major depression: Fact or artefact? *Journal of the American Academy of Child and Adolescent Psychiatry*, *33*, 529–539.

Bird, H. R. (1996). Epidemiology of childhood disorders in a cross-cultural context. *Journal of Child Psychology and Psychiatry*, *37*, 35–49.

Bollen, K. (1989). *Structural equations with latent variables*. New York: Wiley.

Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality*. Unpublished Doctoral Dissertation, Rijksuniversiteit Groningen, The Netherlands.

Browne, M. W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Cantwell, D. P. (1996). Classification of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, *37*, 3–12.

Caron, C., & Rutter, M. (1991). Comorbidity in child psychopathology: Concepts, issues, and research strategies. *Journal of Child Psychology and Psychiatry*, *32*, 1063–1080.

Cattell, R. B. (1973). *Factor Analysis: An introduction and manual for the psychologist and social sciences*. Westport, CT: Greenwood Press.

Cattell, R. B., & Dickman, K. (1962). A dynamic model of physical influence demonstrating the necessity of oblique simple structure. *Psychological Bulletin*, *59*, 389–400.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Dedrick, R. F., Greenbaum, P. E., Friedman, R. M., Wetherington, C. M., & Knoff, H. M. (1997). Testing the structure of the Child Behavior Checklist/4–18 using confirmatory factor analysis. *Educational and Psychological Measurement*, *57*, 306–313.

De Groot, A., Koot, H. M., & Verhulst, F. C. (1994). Cross-cultural generalizability of the Child Behavior Checklist cross-informant syndromes. *Psychological Assessment*, *6*, 225–230.

De Groot, A., Koot, H. M., & Verhulst, F. C. (1995). Cross-cultural generalizability of the Youth Self Report and Teacher's Report Form cross-informant syndromes. *Journal of Abnormal Child Psychology*, *24*, 651–664.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Epkins, C. C. (1994). Peer ratings of depression, anxiety, and aggression in inpatient and elementary school children: Rating biases and influence of rater's self-reported depression, anxiety, and aggression. *Journal of Abnormal Child Psychology*, *22*, 611–640.

Erol, N., Arslan, L. B., & Akçakin, M. (1995). The adaptation and standardisation of the Child Behavior Checklist among 6–18-year-old Turkish children. In J. A. Sergeant (Ed.). *Eunethydis; European approaches to hyperkinetic disorder* (pp. 97–113). Zurich, Switzerland: Fotorotar.

Farrington, D. P., & Loeber, R. (in press). Measures of association: The benefits of dichotomization. *International Journal of Methods in Psychiatric Research*.

Fonseca, A. C., Simões, A., Rebelo, J. A., Ferreira, J. A., Cardoso, F., & Temudo, P. (1995). In J. A. Sergeant (Ed.). *Eunethydis; European approaches to hyperkinetic disorder* (pp. 115–129). Zurich, Switzerland: Fotorotar.

Frick, P. J., Lahey, B. B., Applegate, B., Kerdyck, L., Ollendick, T., Hynd, G. W., Garfinkel, B., Greenhill, L., Biederman, J., Barkley, R. A., McBurnett, K., Newcorn, J., & Waldman, I. D. (1994). DSM-IV field trials for the disruptive behavior disorders: Symptom utility estimates. *Journal of the American Academy of Child and Adolescent Psychiatry*, *33*, 529–539.

Goyette, C. H., Conners, C. K., & Ulrich, R. F. (1978). Normative data on revised Conners parents and teacher rating scales. *Journal of Abnormal Child Psychology*, *6*, 221–236.

Hartman, C. A., Roussos, A. C., Richardson, C., Karandanos, G., Mahaira, O., Zoumbou, V., Kiprianos, S., Berati, S., Georgiadis, D., Lazaratou, H., Karayiannis, D., Papatheofilou, R., & Tassi, M. (1995). Characteristics of the CBCL and the TRF in a Greek population sample. In J. A. Sergeant (Ed.). *Eunethydis; European approaches to hyperkinetic disorder* (pp. 83–95). Zurich, Switzerland: Fotorotar.

Hox, J. (1998). *Simulcat: Software for simulation of categorical data*. Amsterdam: TT Publicaties.

Hox, J., & Hartman, C. A. (1999a). When a poor model meets bad data. In J. J. Hox & E. D. de Leeuw (Eds.)., *Assumptions, robustness, and estimation methods in multivariate modeling* (pp. 141–150). Amsterdam: TT Publicaties.

Hox, J., & Hartman, C. A. (1999b)). *Bootstrapping for fit in structural equation models with non-normal data*. Manuscript in preparation.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351–362.

Jackson, D. N. (1971). The dynamics of structured personality tests. *Psychological Review*, *78*, 229–248.

Jackson, D. N. (1973). Structural personality assessment. In B. B. Wolman (Ed.), *Handbook of general psychology* (pp. 775–792). Englewood Cliffs. NJ: Prentice Hall.

Jöreskog, K. G. (1979a). Basic ideas of factor and component analysis. In K. G. Jöreskog & D. Sörbom (Eds.). *Advances in factor analysis and structural equation models* (pp. 5–20). Cambridge: Abt Books.

Jöreskog, K. G. (1979b). A general approach to confirmatory Maximum Likelihood factor analysis, with addendum. In K. G. Jöreskog & D. Sörbom (Eds.). *Advances in factor analysis and structural equation models* (pp. 21–43). Cambridge: Abt Books.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*, 381–389.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7 User's Reference Guide*. Chicago, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1993a). *New features in PRELIS 2*. Chicago, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1993b). *New features in LISREL 8*. Chicago, IL: Scientific Software International.

Kamphaus, R. W., & Frick, P. J. (1996). *Clinical assessment of child and adolescent personality and behavior*. Boston, MA: Allyn & Bacon.

Kasius, M. C., Ferdinand, R. F., van den Berg, H., & Verhulst, F. (1997). Associations between different diagnostic approaches for child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, *38*, 625–632.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112*, 165–172.

Kovacs, M., & Pollock, M. (1995). Bipolar disorder and comorbid conduct disorder in childhood and adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, *34*, 715–723.

Lachar, D. (1998). Observations of parents, teachers, and children: Contributions to the objective multidimensional assessment of youth. In A. S. Bellack, M. Hersen, & C. R. Reynolds (Eds.), *Comprehensive clinical psychology*, *Vol. 4*. Amsterdam: Elsevier.

Lahey, B. B., Applegate, B., Barkley, R. A., Garfinkel, B., McBurnett, K., Kerdyk, L., Greenhill, L., Hynd, G. W., Frick, P. J., Newcorn, J., Biederman, J., Ollendick, T., Hart, E. L., Perex, D., Waldman, I., & Schaffer, D. (1994). DSM-IV field trials for oppositional defiant disorder and conduct disorder in children and adolescents. *American Journal of Psychiatry*, *151*, 1163–1171.

Lahey, B. B., Applegate, B., McBurnett, K., Biederman, J., Greenhill, L., Hynd, G. W., Barkley, R. A., Newcorn, J., Johnson, P., Richter, J., Garfinkel, B., Kerdyk, L., Frick, P. J., Ollendick, T., Perex, D., Hart, E. L., Waldman, I., & Schaffer, D. (1994). DSM-IV field trials for attention deficit hyperactivity disorder in children and adolescents. *American Journal of Psychiatry*, *151*, 1673–1685.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & Van Kammen, W. B. (1998). *Antisocial behavior and mental health problems: Explanatory factors in childhood and adolescence*. Mahwah, NJ: Lawrence Erlbaum.

Loeber, R., & Keenan, K. (1994). Interaction between conduct disorder and its comorbid conditions: Effects of age and gender. *Clinical Psychology Review*, *14*, 497–523.

Loeber, R., Russo, M. F., Stouthamer-Loeber, M., & Lahey, B. B. (1994). Internalizing problems and their relation to the development of disruptive behaviors in adolescence. Special Issue: Affective processes in adolescence. *Journal of Research on Adolescence*, *4*, 615–637.

Macmann, G. M., Barnett, D. W., Burd, S. A., Jones, T., LeBuffe, P. A., O'Malley, D., Shade, D. B., & Wright, A. (1992). Construct validity of the Child Behavior Checklist: Effects of item overlap on second-order factor structure. *Psychological Assessment*, *4*, 113–116.

Marsh, H., Hau, K. T., Balla, J., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in factor analysis. *Multivariate Behavioral Research*, *33*, 181–220.

McConaughy, S. H., & Skiba, R. J. (1993). Comorbidity of externalizing and internalizing problems. *School Psychology Review*, *22*, 421–436.

Millon, T. (1991). Classification in psychopathology: Rationale, alternatives, and standards. *Journal of Abnormal Psychology*, *100*, 245–261.

Muthén, B. O. (1989). Dichotomous factor analysis of symptom data. *Sociological Methods and Research*, *18*, 19–65.

Muthén, B. O., Hasin, D., & Wisnicki, K. S. (1993). Factor analysis of ICD-10 symptom items in the 1988 National Health Interview Survey on Alcohol Dependence. *Addiction*, *88*, 1071–1077.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171–189.

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19–30.

Nøvik, T. S., & Zeiner, P. (1995). The CBCL and Parental Account of Children's Symptoms (PACS) diagnosis of hyperactivity. In J. A. Sergeant (Ed.)., *Eunethydis; European approaches to hyperkinetic disorder* (pp. 161–171). Zurich, Switzerland: Fotorotar.

Olsson, U. (1979). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research*, *14*, 485–500.

Pliszka, S. (1992). Comorbidity of attention-deficit hyperactivity disorder and overanxious disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *31*, 197–203.

Pulakos, E. D., Schmitt, N., & Ostroff, C. (1986). A warning about the use of a standard deviation across dimensions within rates to measure halo. *Journal of Applied Psychology*, *71*, 29–32.

Quay, H. C. (1986a). Classification. In H. C. Quay & J. S. Werry (Eds.), *Psychopathological disorders of childhood* (3rd ed., pp. 1–34). New York: Wiley.

Quay, H. C. (1986b). A critical analysis of DSM-III as a taxonomy of psychopathology in childhood and adolescence. In T. Millon & G. Klerman (Eds.), *Contemporary directions in psychopathology* (pp. 151–165). New York: Guilford Press.

Rigdon, E. E., & Ferguson, C. E. (1991). The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data. *Journal of Marketing Research*, *28*, 491–497.

Rutter, M., Kebovici, S., Eisenberg, L., Sneznevskij, A. V., Sadoun, R., Brooke, E., & Lin, T. Y. (1969). A triaxial classification of mental disorders in childhood. *Journal of Child Psychology and Psychiatry*, *10*, 41–61.

Rutter, M., & Pickles, A. (1990). Improving the quality of psychiatric data: Classification, cause, and course. In D. Magnusson & L. R. Bergman (Eds.), *Data quality in longitudinal research* (pp. 32–57). Cambridge: Cambridge University Press.

Song, L., Singh, J., & Singer, M. (1994). The Youth Self-Report inventory: A study of its measurement fidelity. *Psychological Assessment*, *6*, 236–245.

Sonuga-Barke, E. (1998). Categorical models of childhood disorder: A conceptual and empirical analysis. *Journal of Child Psychology and Psychiatry*, *39*, 115–133.

Spiker, D., Kraemer, H. C., Constantine, N. A., & Bryant, D. (1992). Reliability and validity of behavior problem checklists as measures of stable traits in low birthweight, premature preschoolers. *Child Development*, *63*, 1481–1496.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Swanson, J. H., Sergeant, J. A., Taylor, E., Sonuga-Barke, E. J. S., Jensen, P. S., & Cantwell, D. P. (1998). Attention-deficit hyperactivity disorder and hyperkinetic disorder. *The Lancet*, *351*, 429–433.

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, *42*, 233–239.

Van den Oord, E. J. (1993). *A genetic study of problem behaviors in children*. Unpublished Doctoral Dissertation, Erasmus Universiteit Rotterdam, The Netherlands.

Van den Oord, E. J., Boomsma, D. I., & Verhulst, F. C. (1994). A study of problem behavior in 10–15-year-old biologically related and unrelated international adoptees. *Behavior Genetics*, *24*, 193–205.

Verhulst, F. C. (1995). The cross-cultural generalizability of the CBCL. In J. A. Sergeant (Ed.), *Eunethydis; European approaches to hyperkinetic disorder* (pp. 161–171). Zurich, Switzerland: Fotorotar.

Verhulst, F. C., & Van der Ende, J. (1992). Six year stability of parent reported problem behavior in an epidemiological sample. *Journal of Abnormal Child Psychology*, *20*, 595–610.

Volkmar, F. R., & Schwab-Stone, M. (1996). Annotation: Childhood disorders in DSM-IV. *Journal of Child Psychology and Psychiatry*, *37*, 779–784.

Waldman, I. D., Lilienfield, S. O., & Lahey, B. B. (1995). Toward construct validity in the childhood disruptive behavior disorders: Classification and diagnosis in DSM-IV and beyond. *Advances in Clinical Child Psychology*, *17*, 323–363.

World Health Organisation. (1967). *International classification of diseases: Manual of the international statistical classification of diseases, injuries and causes of death* (8th rev.). Geneva: Author.

World Health Organisation. (1978). *Mental disorders: Glossary and guide to their classification in accordance with the ninth revision of the international classification of diseases*. Geneva: Author.

World Health Organisation. (1992). *ICD-10: The ICD-10 classification of mental and behavioral disorders: Clinical description and diagnostic guidelines*. Geneva: Author.

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. Scott Long (Eds.), *Testing structural equation models* (pp. 257–293). Newbury Park, CA: Sage.

Wothke, W. (1996). Models for multitrait-multimethod matrix analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 7–56). Mahwah, NJ: Lawrence Erlbaum.

Zilber, N., Auerbach, J., & Lerner, Y. (1994). Israeli norms for the Achenbach Child Behavior Checklist: Comparison of clinically referred and non-referred children. *Israeli Journal of Psychiatry and Related Sciences*, *31*, 5–12.

Zoccolillo, M. (1992). Co-occurrence of conduct disorder and its adult outcomes with depressive and anxiety disorders: A review. *Journal of the American Academy of Child and Adolescent Psychiatry*, *31*, 547–556.

Manuscript accepted 1 April 1999